DOCUMENT RESUME

ED 160 097 IR 006 345

AUTHOR Sutton, John F.; Black, Craig C.
TITLE Data in Systematics Collections. Research Report No. 6.
INSTITUTION Margaret Woodbury Strong Museum, Rochester, N.Y. Museum Data Bank Committee.
PUB DATE Jun 75
NOTE 8p.

EDRS PRICE MF-$0.83 HC-$1.67 Plus Postage.
DESCRIPTORS *Data Collection; *History; *Information Needs; Information Networks; Management; *Museums; *Natural Sciences; *Organization; Problems; Taxonomy
IDENTIFIERS *Data Management

ABSTRACT

Both the collections in natural history museums, and the methods used to organize information about them vary extensively but some data elements are common to all collections, i.e., institution, catalog number, locality, date, preservation, and taxonomy. At this point similarity ends and with the next complexity level, each discipline must decide what data need to be added in order to make their specimens more meaningful; several societies have already set up minimum data sets to be met by all collections in their discipline. Concerns in this field are leaning today toward collection philosophy and management--what should be included in collections and how should they be organized to adequately answer the researcher's questions? Suggestions include combining basic data with other relevant data for each discipline and then finding new methods of organization and manipulation of these collections for increased question-answering capabilities. Computerized inventory lists are suggested as one method, with eventual national information networks for each discipline also being utilized. (Author/MBR)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED160097

U.S. DEPARTMENT OF HEALTH,
EDUCATION & WELFARE
NATIONAL INSTITUTE OF
EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

MUSEUM DATA BANK RESEARCH REPORT

Number 6

Data in Systematics Collections

---

by
John F. Sutton and Craig C. Black

June 1975

This report, or any portion thereof, may be reproduced for any non-commericial purpose provided appropriate credit is given to the authors and the Museum Data Bank Committee as publisher.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

R. G. Chenhall

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC) AND USERS OF THE ERIC SYSTEM."

Museum Data Bank Committee
Robert G. Chenhall, Chairman

Margaret Woodbury Strong Museum
700 Allen Creek Road
Rochester, N.Y.

IR006 345

The majority of natural history museums which exist today contain collections of plants and animals in enormous numbers (Steere 1971). When these collections are thought of as a single resource, the number of specimens involved is truly staggering. The world's herbaria alone house over 200 million specimens of plants (Shetler 1974), and recent mammals in collections across the United States number some 2.5 million specimens (Anderson et al. 1974). These collections represent a rich storehouse of information, but this resource is all too often almost impossible to access. The method of organization of these systematic collections and of this stored information varies from institution to institution (Raven 1974) making it difficult to obtain a single reference datum across the breadth of these collections and almost impossible to recover a series of correlated or dependent data sets from the total collection resource. Is it possible that these collections, which represent all disciplines of natural history, have data elements in common? The Council on Standards of the Association of Systematics Collections believes they do (report in preparation). If this is so, where does one start when looking for common data elements or sets across such a vast and complex array of biological material.

As a starting point, the institutions themselves provide the first piece of common data; that is, the majority of these specimens are in the care of institutions. Each institution provides a theoretical subset of all institutions, and each collection within an institution is a theoretical subset of the total collections. We can extend this one step further by saying that each specimen or lot is a subset of a collection. This progression from the general to the specific provides initial access to the total body of collection information by identifying an institution, an institutional collection and an individual specimen (or lot) generally referenced by some type of numerical system, either by specimen, by lot , or location in the collection. Every specimen is stored somewhere and some record of its existence must be kept if that specimen is to be of any future value to the individual researcher, the institution, or the community at large.

All specimens preserved in collections today must have existed at some point in time. They are entities which were alive and occupied some point or area on or in the surface of the earth. This suggests two, or possibly three, items of data which are likely to be common among the wealth of data found in our natural history collections. The first of these is the time, usually expressed as a calendar date, when the specimen was collected. This date alone encompasses much information, especially when we are dealing with animals which have seasonal migration patterns or molt patterns. A second item of data, usually associated with the date of collection of the specimen, is the place in space from which the specimen was taken. The locality is basic to the nature of the specimen. Without locality data, a specimen is usually not worth the time, effort and expense involved in maintaining it in a collection. Maintenance of specimens is a particularly important and expensive part of natural history collections. It has been estimated that it costs around $10 to place a specimen in the collection of mammals in the American Museum of Natural History and about 24¢ per year to maintain it

thereafter (Anderson 1973). In many collections, condition of the specimen, including the type of preservation, is recorded. The type of preservation can be an especially important item of information as it determines various other data elements available to researchers who may want to study skeletal material, serum, tissues, karyotypes, etc.

The final component which can be assigned to the realm of common data involves taxonomic identifier. Linnaeus established a system of nomenclature which is followed by systematists the world over and which pertains to all biological materials. This nomenclature and the system of classification which has evolved from it provide a general reference system for all of biology and an entry point for all who need and are concerned with access to information on plants and animals whether in the context of research or public service. The taxononmic data may vary in its recorded completeness from one collection to another or one discipline to another, but almost invariably, some taxonomic delimiter is assigned to each specimen in a collection. The taxonomic data in any and all collections is inherently the same although one might never realize this fact when comparing various museum catalogs.

These six basic categories of data, institution, catalog (or lot) number, locality, date, preservation, and taxonomy are or should be associated with every specimen or lot worthy of preservation in natural history collections today. Without this basic set of data, the specimen is virtually useless to any researcher doing systematic work on any group of plants or animals. Assuming that this encompasses the only basic data common to all specimens across the spectrum of systematics collections, can we assume that there is no necessity to have a greater amount of data? What are the requirements of each discipline? What do individuals doing research in a particular group consider as basic?

At this point, similarity between collections of natural history specimens ends. At this next level of data complexity, each discipline must decide what data needs to be added to these basic fields in order to make the specimen more meaningful to them on a discipline level. Several national societies have already adopted minium data sets to be met by all collections in their discipline in order to provide an adequate data base from which systematic biologists can work. The American Society of Mammalogists has established and distributed (Anderson et al. 1974) a list of basic data which must be kept if the collection is to be recognized as a national resource for studies in mammalogy. Other societies, such as the Society of Vertebrate Paleontology, have voted to accept a set of data conventions (Black et al. 1973) as guidelines to data capture but have not made these guidelines mandatory as the mammalogists have done.

There is more and more concern today that systematists take a critical look at their collecting philosophy. The emphasis is slowly shifting from the vacuum cleaner philosophy of collecting to an approach which is predicated on collecting in order to provide specific information for specific problem solving. There is a marked increase in emphasis on quality of specimens and data as well as a retreat from traditional methods of collecting and preserving specimens in order that the research needs of a wider body of scientists may be served. Mammalogists today no longer limit themselves to skin and skull collecting, as was once traditional, but now collect whole

animals that are then preserved in a variety of ways, including nitrogen stored frozen tissue, serum and whole blood samples, whole body preservation in alcohol, etc. These types of collections offer information for types of research in areas other than that of classical systematics. Other specimens may be preserved as whole skeletons which provide additional information often not previously retained and which provide comparative osteological collections for the paleontologist and archaeologist. In this one example taken from mammalogy, we can see that systematic collections may contribute information to a number of disciplines simply by shifting the emphasis slightly and collecting a wider variety of samples preserved in different ways. The traditional skin and skull collections would have been of little use to the paleontologist looking for post-cranial comparative material or to the histologist interested in comparing tissues of clearly related taxa. As a result of these changes in collecting techniques, the systematic collections of mammals are no longer storehouses of information which relate only to species and their distribution but they have become invaluable resources for other disciplines.

This example raises some disturbing questions concerning the amount of data contained in and the usefulness of our systematic collections as they exist today. How does one know if the material he is seeking actually exists in any collections today or further whether these collections have the potential for answering our questions? Do our systematic collections represent a resource which can be tapped by those outside the limited arena of systematic biology? It occurs to us that the basic problem is not what questions can we answer today, but what can we ask and what should we be asking of this information resource. The questions which are answerable with the least amount of effort involve queries about who was where, when and in what numbers. This type of question addressed to any mammal collection as "How many Peromyscus leucopus do you have from Miller's Creek, Montana and collected by John Jones" can be answered very readily by a curator going to his collection of systematically stored specimens, each arranged geographically within each taxon. The curator enters the collection through the data item Peromyscus leucopus, looking for Miller's Creek, Montana and then counts the number of specimens collected by John Jones. At this level of enquiry, a collection of recent mammals curated in the traditional sense can be very useful. However, if questions are more general or more complex, the search capability rapidly deteriorates and vast amounts of curatorial time are consumed. The average query takes from one to ten hours of curatorial time to answer in an average size collection of mammals (Anderson et al., 1974). Systematics collections appear to be organized primarily to answer the basic question of "Do you have any specimens of Xux yus from a particular locality."

Considering only the basic data set forth at the beginning of this paper, there are many other questions which can be asked of a systematic collection but which cannot be easily answered as the collection information is presently organized. Combining these basic data with other data collected as relevant for each discipline, the number of questions which may be answered increases exponentially. Standard cataloging techniques, including card files, ledger books and the taxonomic arrangement of collections themselves, thus become giant pools of data which are capable of answering only a limited number of questions. With the increasing numbers of specimens

added to these collections each year (best estimate of growth is one to two percent per year (Irwin et al. 1973)), these pools quickly tend to become opaque and incapable of easily providing answers to even the simplest questions.

We must look for new methods of organizing and manipulating these large volumes of data so that more questions and new questions can be asked and answered in a reasonable amount of time. We believe that systematics collections have reached a point in their development, at least in the United States, where justification for their continuing support and maintenance must be made not only on the basis of their research contributions, but also on the basis of their service in areas of public concern, such as environmental assessment and monitoring. Systematics collections can no longer be totally justified on the basic premise that they are (ideally) repositories for all elements of the biological world and that they service systematic biologists.

New insights need to be gained; new questions must be formulated. What information are we acquiring and what are we losing from the animals and plants we are collecting today? Can we improve collecting techniques to provide more information on the biological whole, the ecosystem, rather than focusing, as we do in most cases, on a group of animals or plants? Those concerned with the making and maintenance of systematics collections must change, or more properly expand, their outlook so that this national resource is capable of providing information on the changing biological world. Populations of animals are not static but are extremely dynamic, and future collections must look at biological units in this light. Collections made now and data generated from these collections to facilitate an environmental statement cannot be treated as providing the final word. Collections must be made at regular intervals in the future and we must be able to readily compare these with the earlier baseline data in order to assess the change in a community brought about by naturally occurring phenomena and also by the man-made changes which perhaps first spurred the study of an area. It is this continual monitoring of environmental situations which will, in time, give man some degree of ecological predictability, far removed from our poor attempts today. This monitoring must be carried out by systematists and ecologists and will only be possible through the utilization of the latest electronic data manipulation techniques.

This brings us back to the problem of data management and organization. It is imperative that the wealth of information which already exists in systematics collections, as well as new data accruing through collection growth be handled in a more efficient manner than it is presently. The term "computer" strikes fear into the cataloging heart of many traditional systematic biologists, but it is only through electronic data handling systems that any progress will be made in controlling the quality and quantity of data provided by natural history collections. Certainly, when one looks at the 200 million specimens contained in the world's herbaria and talks of data banking all of the associated data, the task seems insurmountable. In fact, in any discipline, the problems would be the same; retroactive cataloging will require more personnel and machine time than any institution or discipline can presently afford. Looking

at this problem from another viewpoint, however, will we gain anything by continuing to collect biological specimens in the same manner as before? Obviously, the system which has failed to produce usable data on a wide front in the past will not heal itself and become more useful in the future. As users and managers of systematic collections, now is the time to start recording our data in a machine readable form. As time and personnel permit, we may then pick up existing specimen information as it becomes pertinent to specific projects.

Even if all existing data were present in some large computer bank in the sky, it is highly unlikely that this data could be manipulated to generate any meaningful environmental statement about any specific area. It is probable, however, that data does exist in some collections for any given area and these data may be helpful in starting environmental investigations by outlining what has been done in the past and what type of collecting effort needs to be made in the future. It seems likely that an inventory list of existing specimens and localities would be useful as a starting point for both researchers and general users. However, complete collections from a specific site are not likely to be found because of collecting methods employed in the past.

The possibility of generating inventory lists does exist but may not be realized in the immediate future. The starting point for any inventory list has to be the individual specimen with its set of basic data which provides general statements about that specimen. As the trend toward computerization of institutional collections grows, as seems probable, the basis for a larger data banking system becomes more evident. Such a system should probably not be based on specimen identifier entry, but perhaps on taxon-locality entry in which each institution would generate a list containing each unique taxon-locality combination which could be permuted from its existing computerized files. To be most useful, this would have to be done discipline by discipline due to the variability in collecting methods from one discipline to another. Sheer masses of data would also mitigate against a single file containing taxon-locality information for all systematics collections.

If any sort of computerized inventory list is to be possible, care must be taken to design a system which would not fall into the many pitfalls which were so clearly enunciated by Shetler (1974). Data banking, with all its implications, indeed has myths, as Shetler has said (1974), but a realistic approach to implementation, perhaps one on a regional basis, must be found if it is ever to become a reality. We must also critically define the uses which such a data bank is to serve before we begin its construction.

As Shetler (1974) also points out, we have the machines now which make such a inventory system technically feasible; however, there is still an enormous need for manpower, especially for input of data, and this is a major block in the road towards implementation. It is for these reasons that we believe that any sort of national data bank of all systematics information is still a fair distance in the future. With the current trends toward computerization of data in many of the larger

collections in the United States, local data bases will eventually be built to a point where it may be possible to merge particular data blocks and combine them into a national system organized most probably discipline by discipline.

The possibility of a national system such as this must be based in need and realistic service. This service is dependent upon and will define the data put into the system. We believe that each data bank must include at least the basic data elements which we have considered here. Without these basic blocks there can be no link to what was done in the past and what needs to be done in the future. Without these data at least, we cannot hope to monitor the environmental changes which are occurring daily and which affect every aspect of our daily lives. We believe that the biological world must be studied with renewed interest and that collections of animals and plants are basic to this study. We must know our environment and we must establish priorities as to how and why we modify it.

REFERENCES CITED

Anderson, S.

1973 It costs more to store a whale than a mouse: libraries, collections, and the cost of knowledge. Curator 16(1): 30-44.

Anderson, S., et al.

1974 Report and recommendations. Advisory committee for systematic resources in mammalogy. American Society of Mammalogists.

Black, Craig C., et al.

1973 Data conventions for specimen record files. Society of Vertebrate Paleontology News Bulletin. 97:61-69.

Irwin, Howard S., et al.

1973 America's systematics collections: a national plan. Association of Svstematics Collections.

Raven, Peter H.

1974 Summary remarks. Applications of systematics collections: the environment. Association of Systematics Collections Symposium

Shetler, Stanwyn G.

1974 Demythologizing biological data banking. Taxon 23(1):71-100.

Steere, William C.

1971 The systematic biology collections of the United States: an essential resource, Part 11. The great collections: statistical information. Report to National Science Foundation.